Particip-AI: A Democratic Surveying Framework for Anticipating Future AI Use Cases, Harms and Benefits

Jimin Mun[1], **Liwei Jiang**[2, 3], **Jenny Liang**[1], **Inyoung Chung**[2], **Nicole DeCario**[3], **Yejin Choi**[2, 3], **Tadayoshi Kohno**[2], **Maarten Sap**[2, 3]
[1]Carnegie Mellon University, Pittsburgh, PA, USA
[2]University of Washington, Seattle, WA, USA
[3]Allen Institute for AI, Seattle, WA, USA
jmun@andrew.cmu.edu

Abstract

General purpose AI, such as ChatGPT, seems to have lowered the barriers for the public to use AI and harness its power. However, the governance and development of AI still remain in the hands of a few, and the pace of development is accelerating without a comprehensive assessment of risks. As a first step towards democratic risk assessment and design of general purpose AI, we introduce PARTICIP-AI, a carefully designed framework for laypeople to speculate and assess AI use cases and their impacts. Our framework allows us to study more nuanced and detailed public opinions on AI through collecting use cases, surfacing diverse harms through risk assessment under alternate scenarios (i.e., developing and not developing a use case), and illuminating tensions over AI development through making a concluding choice on its development. To showcase the promise of our framework towards informing democratic AI development, we run a medium-scale study with inputs from 295 demographically diverse participants. Our analyses show that participants' responses emphasize applications for personal life and society, contrasting with most current AI development's business focus. We also surface diverse set of envisioned harms such as distrust in AI and institutions, complementary to those defined by experts. Furthermore, we found that perceived impact of *not* developing use cases significantly predicted participants' judgements of whether AI use cases should be developed, and highlighted lay users' concerns of techno-solutionism. We conclude with a discussion on how frameworks like PARTICIP-AI can further guide democratic AI development and governance.

1 Introduction

In response to rapid adoption of AI and expansion of its application areas, calls for more democratic and comprehensive risk assessment of AI are growing (Maslej et al. 2023; Tracy 2023; Bengio et al. 2023). Yet, these pose several challenges. On one hand, the current assessment frameworks and development decisions have largely been guided by experts (Solaiman et al. 2023; Weidinger et al. 2023; Barrett et al. 2023), overlooking the broadening impact of AI to and its widening usage by everyday, non-expert public (Viswanathan et al. 2023; Center 2023). On the other hand, participatory frameworks for AI have been adopted in many specific domains (Corbett, Denton, and Erete 2023; Friedman, Kahn Jr., and Borning 2008; Friedman and Hendry 2012), but the flexible and instructable nature of general-purpose AI (Zoph et al. 2022) requires large-scale and diverse participation in *anticipating* use cases of AI in addition to use case development, to comprehensively evaluate its possible impacts. Thus, tackling these challenges is the key towards a more open and less centralized decision-making (Widder, West, and Whittaker 2023; Brynjolfsson 2023) around AI design and governance.

A necessary first step towards this goal is to build a framework for the non-expert public to share opinions and express critical assessments on AI. Such a framework must be centered around concrete use cases (Trustible 2023; Parliament 2023), since only discussing high-level regulation of general purpose models leads to rules that are too vague to operationalize (Rao and Kreps 2023). Moreover, the framework should allow the public to consider the alternate reality associated with an AI use case, considering both its development and *non*-development, a contrastive perspective often missing in technology assessment (Forlano and Halpern 2023). Finally, to widely speculate the future of AI, the use cases and development scenarios should cover futuristic, e.g., AI in 5–10 years, use cases and their impact as well.

Towards this goal, we introduce PARTICIP-AI, a framework to gather detailed and nuanced public opinion on AI based on current and future use cases and their impact, inspired by speculative design practices (Hohendanner et al. 2023; Balaram, Greenham, and Leonard 2018) and consequentialist ethics modeling (Kohno, Acar, and Loh 2023; Card and Smith 2020). Our framework proposes a four-step process that asks participants to brainstorm possible use cases, imagine and rate their harms and benefits under two alternate scenarios of developing and not developing, and finally, make a choice on the development of the use cases.

To show the feasibility of our framework, we conduct an online survey with 295 demographically diverse, US-based participants and analyzed their responses, qualitatively and quantitatively, to answer the following research questions.

RQ1 To explore overlooked AI development directions and help guide equitable progress through public input, we ask: ***what current and future use cases of AI are in the public's imagination?***

RQ2 As many seemingly beneficial use cases of AI have problems of dual use and failures, we analyze: ***what are the harms and benefits of the use cases?***

RQ3 To gather input beyond technological determinism (Littman et al. 2022), we examine: ***what are the harms and benefits of not developing certain applications of AI?***

RQ4 To study people's decision processes and conflicting values, we explore: ***what creates tension between developing and not developing the applications?***

Our results surface a wide array of anticipated use cases (RQ1), which highlight common themes of interest in improving personal, everyday life, showing diverse interests to augment life through AI and emphasis on the value of AI in making societal impact towards betterment of society as a whole. We find that our participants surface set of harms complementary to taxonomies created by experts (RQ2), for example, raising issues of distrust in institutions and highlighting the need for regulation to protect mental health. Moreover, our findings uncover a set of benefits and harms associated with not developing a use case and highlight a tension in AI's impact on human potential (RQ3). Finally, we find that level of benefits and harms of *not* developing a use case is significantly more correlated with decisions of development, compared to that of developing (RQ4).

To summarize our contribution, we (1) propose a PARTICIP-AI, a novel framework to assess AI use cases and their impact, developed with insights from various field of AI, computer security, and philosophy. We (2) conduct a survey with lay users to showcase PARTICIP-AI's usability. We (3) present results from synthesizing themes such as interest in use cases that emphasize equitable progress through enhancing everyday life and solving societal issues, harm types complementary to the expert-generated, various impacts of not developing a use case, and tensions over value of human work. Finally, we (4) conclude with a discussion on direction of AI development to reflect diverse goals and needs, risks of AI and ways to address regulatory gaps, and tensions over development and techno-solutionism. Our work highlights the promise and importance of including lay publics and diverse voices into the future of AI design and governance.[1][2]

2 Related Work

Participation in AI. While the rapidly growing deployment of AI systems across many sectors has called for meaningful participation (Costanza-Chock 2018; Delgado et al. 2023a; Queerinai et al. 2023; Pasquale, Malgieri, and Pasquale 2021; Suresh et al. 2022), exploration of such approaches has lagged behind (Birhane et al. 2022; Bergman et al. 2023; Durmus et al. 2023; Bender et al. 2021), especially in large-scale AI models (Birhane et al. 2022). On one hand, previous efforts span "data labor" for model optimization such as annotation and feedback (Miceli and Posada 2022; Bai et al. 2022a), enabling human inputs at granular, instance-level (e.g., human annotation or feedback; OpenAI 2023; Bai et al. 2022a; Christiano et al. 2023) or at limited stages of AI pipeline (e.g., representative evaluation; Bergman et al. 2023) largely for existing AI systems. On the other hand, many previous works focused on a broad, principles-level, participation including constitutional AI (Bai et al. 2022b; Anthropic 2023), citizen juries (Balaram, Greenham, and Leonard 2018; van der Veer et al. 2021), public AI policy insights (Reeve, Colom, and Modhvadia 2023; NIST 2023), and community collectives (Nekoto et al. 2020; Queerinai et al. 2023).

Unlike previous works, our work targets the middle-ground by facilitating public assessment of potential real-world AI applications across domains, aligning with recent legislation advocating more application-based approaches (Parliament 2023). Our work takes inspiration from design futuring (Fry 2009; Kozubaev et al. 2020), including speculative design and design fiction, which focuses on non-linear approaches that seeks to challenge and criticize the status quo, explore alternate scenarios, and (re)envision the future (Hohendanner et al. 2023; Farias, Bendor, and Van Eekelen 2022; Baumann et al. 2016, 2018). Thus, our framework addresses the limitations in previous works through future-looking, use-case oriented questions, which allow for discussion around broader stages of AI pipeline (e.g., usage, design, threats, and opinions of deployment stages) and while we implement the framework as a survey in this work, is not limited to any specific format.

Risk Assessment of AI Applications. AI applications' far-reaching impact and increasing accessibility among lay users (Viswanathan et al. 2023) urges broad, deliberate, and multifaceted assessments of their nuanced and unexpected risks (Lubars and Tan 2019). While many works have developed assessment frameworks of AI risks, most have focused on expert inputs only (Solaiman et al. 2023; Weidinger et al. 2023; Barrett et al. 2023), neglecting the valuable perspectives of end users impacted by AI-related harms, focusing on broader guidelines (e.g., human rights; Prabhakaran et al. 2022), overlooking conflicting values of diverse set of users.

In works that incorporate user inputs for AI risk assessments, there is a noted limitation in accommodating a wide range of diverse and potentially conflicting human values (Weidinger et al. 2023). For example, these works cover limited deployment scenarios (Buçinca et al. 2023) or targets specific stage of AI development pipeline and user group (e.g., tool to help AI developers and researchers in prototyping harms; Wang et al. 2024). To address the challenges of surfacing diverse perspectives and potentially conflicting values, our framework adopts a large-scale participation-based approach with lay-users to anticipate risks associated with current and future AI. Moreover, by soliciting lay users' inputs to assess the potential harms and benefits of both *developing* and *not developing* AI applications our work more comprehensively gathers conflicting interests and values.

3 Methods

In this section, we first introduce PARTICIP-AI framework, including the motivation and scope, and the question choices

[1]See https://github.com/JiMinMun/Particip-AI for all participant responses.

[2]Supplementary material and appendix are at https://arxiv.org/abs/2403.14791.

Question Numbers	Content
Technology Description	
⚙ Tech-X	Imagine an AI technology (let's call it "Tech-X") is developed by tech companies that can follow any instructions to generate new content such as images, human-like language, computer code, etc. To name a few of its capabilities, it can interact with people in a conversational way, write stories, create illustrations and paintings, and answer questions about almost anything.
⚙⚙ Tech-X 10	Now consider a date five to ten years into the future. Imagine a more sophisticated version of Tech-X (let's call it "Tech-X 10"), which can follow any instruction you give it, has expert-level knowledge or even better, can solve problems creatively, can connect to the internet and other devices, and can process and read massive amounts of data or text within seconds.
Sec 1: Use Cases	
⚙ Q1, ⚙⚙ Q4	Do you think a technology like this should exist?
⚙ Q2, ⚙⚙ Q5	How confident are you in the above answer?
⚙ Q3*, ⚙⚙ Q6*	Please write some tasks that {⚙, ⚙⚙} could assist or automate in the text box below.
💼 Q7*	Complete the following sentence by choosing one task from your brainstormed answers that you believe Tech-X or Tech-X 10 will change the most drastically. The task that I think Tech-X / Tech-X 10 would most dramatically change would be in…
Sec 2: Harms and Benefits of Developing	
👍 Q8*, 👎 Q11* & Q14*	How will Tech-X / Tech-X 10 automating or assisting the task you identified {👍, 👎} impact individuals?
👍 Q9*, 👎 Q12* & Q15*	Which groups of people do you think would {👍, 👎} the most from the above {👍, 👎} impacts?
👍 Q10, 👎 Q13 & Q16	How {👍, 👎} would it be if Tech-X / Tech-X 10 was used for the following task and had the above {👍, 👎} impacts?
Sec 3: Harms and Benefits of *Not* Developing	
👎 Q17, 👍 Q20*	Now imagine that Tech-X / Tech-X 10 was never used to automate or assist with {💼}. How will banning or not developing this particular application {👍, 👎} impact individuals?
👎 Q18, 👍 Q21*	Which groups of people do you think would {👍, 👎} the most from the above by banning or not developing this particular application?
👎 Q19, 👍 Q22	How {👍, 👎} would it be if Tech-X / Tech-X 10 was banned or never developed to perform the following task and had the above {👍, 👎} impacts?
Sec 4: Use Case Opinion	
📝 Q23	After thinking about the benefits and harms of the application and the harms of it not being developed, do you think that this application of the technology should or should not be developed?
Q24	How confident are you in the above answer?
Q25	How likely do you think are people to agree that an application of Tech-X / Tech-X 10 that automates or assists with {💼} {📝}?

Table 1: Survey questions in PARTICIP-AI. Questions are summarized due to space constraints. Asterisks (*) denote open-text questions. Within curly brackets are variations such as benefit (👍) or harm (👎) or input from previous questions, e.g., task (💼).

for the survey instrument (§3.1). We then describe our methods for collecting and analyzing lay users' inputs (§3.2).

3.1 PARTICIP-AI Framework

Overview The primary goal of PARTICIP-AI is to effectively elicit the opinions of lay-users on the potential harms and benefits across many near- and far-future AI applications. To create such a framework that addresses pertinent, pivotal questions in AI and to incorporate diverse, bottom-up views, we harness the interdisciplinary expertise of our research team, spanning computer security, public policy, natural language processing (NLP), and AI ethics. We *iteratively* design a survey to reflect our research questions and adopt an *online crowdsourcing platform* for broad and controlled distribution to populations with diverse backgrounds.

To explore how lay users perceive the influences and consequences of *future* AI applications, we prompt users to *imagine* potential use cases of AI and consider speculative harms and benefits of both *developing* and *not developing* such technologies. This fictional inquiry approach (Dindler and Iversen 2007), is inspired by various fields, including design fictions (Bleecker 2022) and threat modeling in computer security (Evtimov et al. 2020). In particular, the alternative scenarios (i.e., *to develop* or *not to develop* a use case) involve choosing between two outcomes, reminiscent of hypothetical dilemmas in moral philosophy (Bostyn, Sevenhant, and Roets 2018).

The option of *developing* an AI use case considers two distinct types of harms: (1) those arising from the *failure or low performance* of AI (Raji et al. 2022), and (2) those resulting from the *malicious misuse* of AI (Pöhler et al. 2024). Finally, acknowledging the distinct real-world impacts of *short-term, near-future*, and *long-term, far-future* AI technologies, PARTICIP-AI guides users to analyze and differentiate the distinct potential harms and benefits presented by AI with varying levels of capabilities.

Question Design Survey questions are shown in Table 1.

Use Cases of AI (RQ1): First, participants are asked to imagine the use cases of two variants of AI, *Tech-X* and *Tech-X 10*. Tech-X, while fictional, describes a technology similar to current generative AI, i.e., instruction-following with generative output. Tech-X 10, on the other hand, describes a technology five to ten years into the future with a focus on its expert-level knowledge and creative problem-solving. For each variant, participants are asked three questions: whether the technology like the one described should exist or not (Q1, Q4; binary), their confidence in that opinion (Q2, Q5; 5-point Likert), and ideas on use cases of the technology (Q3, Q6; free-text). Finally, participants are asked to choose one brainstormed use case that would be changed *most drastically* through AI (Q7). *All subsequent questions ask specifically about the use case chosen in this step.*

Harms and Benefits of Developing (RQ2): Here, participants are asked to anticipate the use case's benefits and the two types of harms (i.e., malicious misuse, failure cases). Regarding the benefits and each type of harm, participants describe their impact (Q8, Q11, Q14; free-text), the most impacted groups (Q9, Q12, Q15; free-text), and the scale of the impact (Q10, Q13, Q16; 8-point Likert[3]).

Harms and Benefits of NOT Developing (RQ3): Next, assuming a hypothetical scenario where the technology is not used for the use case, participants are tasked to describe the impact of potential harms and benefits (Q17, Q20; free-text), most impacted groups (Q18, Q21; free-text), and the scale of impact (Q19, Q22; 8-point Likert[3]).

Use Case Opinions (RQ4): Finally, to understand how participants perceive the permissibility of developing the use case, participants are asked to select whether the application should be developed (Q23; binary), the confidence in that answer (Q24; 5-point Likert), and how likely it would be that others would agree to that opinion (Q25; 8-point Likert[3]).

3.2 Input Collection and Analysis Methodology

We performed targeted recruiting of diverse participants to conduct our survey and a mixture of *quantitative* (for questions with nominal or ordinal scale answers) and *qualitative* analyses (for questions with free-text answers) to extract insights from the collected survey data.

Participant Recruitment We recruited 300 participants on Prolific to conduct the PARTICIP-AI survey.[4] To obtain diverse opinions, we performed targeted recruiting across five different ethnic groups provided by the platform (i.e., Asian, Black, Mixed, Other, and White), and across two age groups (i.e., 18 to 48, 49 to 100) that divides the US adult population in approximately half,[5] resulting in 10 different groups with 30 participants each. These groups were balanced in male and female sex categories. 295 participants responded to the survey. The survey took median 25.2 minutes. Participants were compensated at the rate of $9.67/hr. Our study was approved by an institutional review board. Limitations due to the platform's recruitment, stratification, and categorization methods are discussed further in §5.

Below, we describe a subset of participants' demographics; see Appendix A.2 for a full description. While we recruited equal number of demographically stratified participants, imbalances occurred due to the specification of self-identified racial categories and lack of responses in certain groups. Participants were White (29.8%), Black (21.7%), Asian (20.0%), Other (10.8%), and Mixed (11.9%). Participants largely identified as men (47.8%) and women (47.1%).

Qualitative Analysis We apply qualitative analysis to free-form answers, using concept coding (Saldaña 2021) practices augmented with GPT-4[6] to aid in applying the human-generated codebook on the large-scale survey dataset. These questions include AI use cases (Q3, Q6, Q7), impact of benefits (Q8, Q20), impact of harms (Q11, Q14, Q17), and most impacted groups (Q9, Q12, Q15, Q18, Q21). See Appendix H for the codes and their definition.

Open Coding with Human Annotators With three authors from our research team, we developed codebooks by performing open coding on approximately 25% of the data (80 samples). For questions on tasks and impact of harms and benefits (Q7, Q8, Q11, Q14, Q17, Q20), three authors developed codebooks independently, and then merged them into a shared codebook upon discussion with unanimous agreement. Next, the three authors applied the merged codebooks to all questions of each user sample. Individual low-level codes are grouped into a high-level *theme* during data analysis. In addition, we pre-process brainstormed tasks (Q3, Q6) using GPT-4 to standardize their expressions (see Appendix B.1 for further details). Finally, given the brevity and directness of answers, questions pertaining to groups impacted (Q9, Q12, Q15, Q18, Q21) were coded by a single author. See Appendix B.2 for further details.

Closed Coding with GPT-4 Augmentation Manually coding responses is prohibitive when the sample size is large; frontier language models such as GPT-3 and GPT-4 have shown promise in automatic qualitative coding (Xiao et al. 2023; Matter et al. 2024). Thus, we applied GPT-4 to perform closed coding of the remaining 75% of the samples using the codebooks developed by our research team during the open coding process. We use the first two samples from the held-out data that we manually coded as few-shot examples to prompt GPT-4, and we evaluate human to GPT-4 agreement using the remaining 78 samples. In line with previous research (Xiao et al. 2023), GPT-4 has moderate (0.41-0.60) to substantial (0.61-0.80) agreement (McHugh 2012) with the human annotators.[7] See Appendix B.3 for detailed prompts, settings, and additional metrics.

Quantitative Analysis. Quantitative analysis is conducted on data in nominal or ordinal scales, including opinions on whether a use case should be developed (Q1, Q4, Q23; binary), the scale of impact (Q10, Q13, Q16, Q19, Q22; Likert) and participants' rating of their confidence and anticipated agreement (Q24, Q25; Likert). We aggregate the percentage of responses for opinions, take the mean by theme

[3]Anchored scale to control for individual user interpretations. See Appendix A for details.

[4]https://www.prolific.com/

[5]US Census Data, Accessed on 11/29/2023.

[6]gpt-4-1106-preview

[7]Inter-rater agreement scores based on Cohen's κ (Cohen 1960) on validation set for each question: (Q7; κ=.59, Q8; κ=.51, Q11; κ=.66, Q14; κ=.67, Q17; κ=.62, Q20; κ=.58, Q9; κ=.77, Q12; κ=.85, Q15; κ=.87, Q18; κ=.85, Q21; κ=.82)

for the harms and benefits scales, and conduct an exploratory factor analysis for the effects of harms and benefits on opinions of use case development.[8]

4 Results

In this section, we present and discuss the results of our survey based on PARTICIP-AI framework with 295 participants. Before delving into the specific use cases, participants showed an overall positive attitude towards the general descriptions of the AI technology with 86.1% (Q1) and 85.4% (Q4) responding that Tech-X and Tech-X 10 "Should exist".

4.1 RQ1. Current and Future Use Cases of AI

To answer this question, we analyze the brainstormed use cases for current (Q3) and future technology (Q6), as well as the tasks that would be most drastically impacted by AI (Q7). We grouped the codes (see Appendix C) into the high-level *themes* to assist analyzing the results and highlight general trends (see Table 2): DOMAIN, SUPPORT TYPE, realms of impact (i.e., WORK, PERSONAL LIFE, and SOCIETY), and GOAL OF THE USE CASE.

Current and Future AI Use Cases Participants brainstormed a similar number of use cases across themes for both current (Q3, Tech-X; avg. of 3.5) and future (Q6, Tech-X 10; avg. of 3.4 tasks) technology. Across both questions, participants most commonly mentioned the DOMAIN of use cases (55.0%; Q3, 63.2%; Q6) compared to SUPPORT TYPE, GOAL, or realms of impact. However, use cases differed in their distributions within the theme: notably, those for future technology emphasized domains such as medical (10.5%), education (9.7%), and research (3.0%) whereas those for the current version discussed artistic expression (13.4%), education (11.1%), and translation (8.0%). PERSONAL LIFE applications occured more frequently for Tech-X (45.7%) compared to Tech-X 10 (39.2%). In contrast, tasks surrounding impact to SOCIETY grew most drastically from Tech-X (0.3%) to Tech-X 10 (8.7%), suggesting people's interest in future AI applications to address societal issues. See Table 2 for detailed distribution of the themes and codes.

Participant Selected Use Cases Among tasks described as most revolutionized by AI, DOMAIN of applications was the most common theme (67.8%), covering medical (13.9%), education (10.5%), and research (10.5%) domains. The second most prevalent theme was SUPPORT TYPE (40.7%) containing top use cases related to efficient data analysis (19.0%) and professional consulting service (7.8 %). As in previous questions, PERSONAL LIFE (29.5%) related tasks were discussed more frequently compared to WORK (10.8%) and SOCIETY (9.5%) (see Appendix C.2). Notably, participants selected more use cases that impact *the society* compared to previous questions (see Appendix C.3).

[8]For effects analysis, we converted participants' answers numerically: opinion (-1="Should not be developed", 1="Should be developed"), confidence (-4="Should not be developed" $\times$ "Extremely confident", 4="Should be developed"$\times$"Extremely confident") and perceived agreement (-8="Should not be developed" $\times$ "Highly likely", 8="Should be developed" $\times$ "Highly likely")

4.2 RQ2. Harms and Benefits of the Use Cases

To answer this queston, we analyzed participants' anticipated harms and benefits of their use case (qualitative; Q8, Q11, Q14), groups that could be harmed or benefited the most (qualitative; Q9, Q12, Q15), and the scale of impact (quantitative; Q10, Q13, Q16). Analysis on questions such as anticipated groups affected by developing use case (Q9, Q12, Q15) can be found in Appendix D.2.

Harms of Developing The harms of developing the selected use cases were grouped into ten high-level themes (see Table 3; left). We analyzed harms due to misuses (Q11) and poor performance (Q14) separately. For harms due to *misuses or unintended consequences*, participants most often mentioned SOCIAL AND PSYCHOLOGICAL EFFECT (35.3%) followed by ECONOMIC IMPACT (32.5%). Within SOCIAL AND PSYCHOLOGICAL EFFECT, the most common concerns were manipulation of people (12.9%) (e.g., "*control and manipulate information for human exploitation*" by P114), misinformation (12.5%), and mental harm (12.2%). For harms caused by the poor performance of technology, participants most commonly discussed ECONOMIC IMPACT (33.6%) at the personal and societal level, such as financial disturbance (20.7%) and economic disturbance (9.8%). The second most discussed harm due to failure cases was PHYSICAL EFFECT (29.8%), such as physical harm (23.7%) and negative impact to health and well being (8.5%).

While the two types of harm showed different distributions of themes, their scale of impact was similar. Among all themes, REDUCING PROGRESS (7.50$\pm$0.84; Q13, 6.57$\pm$1.45; Q16) and PHYSICAL (6.70$\pm$1.40; Q13, 6.28$\pm$1.59; Q16) had the biggest scale of impact (see Table 3). While ECONOMIC IMPACT was a frequent theme overall, its perceived impact was lower (5.39$\pm$1.70; Q13, 4.63$\pm$1.75; Q16), especially for poor performance harms.

Benefits of Developing The benefits of selected use cases are grouped into eight themes (see Table 4; right). The most prominent theme was REINVESTING HUMAN CAPITAL (52.5%), within which, personal life efficiency (35.6%) to "*save time effort and energy*" (P19) in personal life was mentioned the most followed by personal growth (16.9%), and reducing mundane work (13.2%). The second most frequent theme was ECONOMIC GAIN (43.7%), such as general efficiency (31.5%) and financial gain (17.6%).

While REINVESTING HUMAN CAPITAL was the most frequently observed benefit, its scale of impact (5.13$\pm$1.86) was lower compared to IMPROVING QUALITY OF SOCIAL LIFE (6.76$\pm$1.17) and IMPROVING QUALITY OF PERSONAL LIFE (6.50$\pm$1.50). This suggests that while AI offers efficiency in reinvesting human capital, the more influential positive impact comes from improving the quality of life.

4.3 RQ3. Harms and Benefits of *Not* Developing Certain Applications of AI

We analyzed participants' answers on harms and benefits of not developing (qualitative; Q17, Q20), groups that could be harmed or benefited the most (qualitative; Q18, Q21), and the scale of impact (quantitative; Q19, Q22) to answer this

Code / THEME	Quote	% responses Q3 1032×	Q6 992×	Q7 295×
DOMAIN		**55.0**	**63.2**	**67.8**
Artistic expression	*"Make abstract art with my dog's picture"* (P13, Q3)	13.4	3.4	5.4
Medical	*"Automating medical research"* (P54, Q6)	4.9	10.5	13.9
Education	*"Teaching me how to code in different languages"* (P104, Q7)	11.1	9.7	10.5
Research	*"Revolutionize scientific research"* (P203, Q6)	3.0	8.5	10.5
Translation	*"Translations while traveling"* (P292, Q3)	8.0	1.3	4.1
SUPPORT TYPE		**39.0**	**39.3**	**40.7**
Efficient data analysis	*"Assist in personalized medicine by analyzing genetic data, medical histories, and current research..."* (P242, Q6)	8.9	17.4	19.0
Professional consulting service	*"Mental health diagnosis and intervention...since...professionals often gets overwhelmed with their work."* (P203, Q7)	2.1	6.3	7.8
Writing assistance	*"Adapting resumes to different job postings"* (P146, Q3)	11.5	3.6	2.7
PERSONAL LIFE		**45.7**	**39.2**	**29.5**
Everyday task automation	*"Summarizing important email content into a list"* (P293, Q3)	25.4	22.1	14.2
Everyday life assistance	*"Assisting me with planning meal ideas that meet my family's dietary needs...[to] take much weight off my mental plate."* (P249, Q7)	17.8	15.6	13.9
GOAL OF THE USE CASE		**26.5**	**25.9**	**33.6**
Personal life productivity	*"Assisting with managing time spent on activities"* (P100, Q3)	8.8	6.0	10.2
Creativity	*"Creating unique entertainment options that cater to individuals and evolve with them over time"* (P265, Q6)	12.4	4.6	6.4
SOCIETY		**0.3**	**8.7**	**9.5**
Societal issues	*"Predictive models that improve health and environment challenges."* (P28, Q7)	0.3	8.7	9.5
WORK		**5.2**	**6.8**	**10.8**
Human labor replacement	*"Replacing humans in customer interaction jobs"* (P282, Q6)	0.8	4.8	8.1
Workplace productivity	*"Generate job reports that would take human hours"* (P86, Q6)	4.5	2.5	4.1
OTHER		**0.8**	**2.2**	**4.4**
New code	*"Find a way for the AI to destroy its own AI self"* (P25, Q6)	0.8	2.0	3.7

Table 2: Tasks (Q3, Q6, Q7): percentage of occurrence for THEME and top few most frequent codes with representative quotes.

question. See Appendix E.2 for analysis on questions about anticipated groups affected by use case (Q18, Q20).

Harms of Not Developing Responses on harms of not developing the use case are grouped into nine high level themes (see Table 5). The most common themes were LIMITING HUMAN POTENTIAL (32.9%) and LOSE INFORMATION AND ACCESSIBILITY TO RESOURCES (26.1%). By not developing the application, it's anticipated that there will be more wasted resources or time (13.2%) or inefficiency (12.2%), e.g., "*where people's lives are being wasted on unfulfilling labor for low pay*" (P110). Another major concern involves losing assistance for the task (11.5%) and losing accessibility to solution and service (9.8%). Unlike the harms of developing (§4.2), 23.4% of answers were categorized as OTHER, within which many answers mentioned there being no harm (16.6%), indicating that harms of not developing often does not exist or is harder to imagine compared to harms of developing (e.g., "*if it never gets develop [sic] we won't know what we are missing out on*").

Regarding the scale of impact (Q19; see Table 5), PHYSICAL EFFECT had the highest perceived impact of harm (5.14±2.62), similar to the scale of impact indicated in harms of developing. Participants also anticipate a high impact of LESS PROGRESS IN SOLVING SOCIETAL ISSUES (4.78±2.19), which, considering previous results, conveys solving societal issues an important beneficial area of AI.

Benefits of Not Developing Benefits of not developing the use cases had seven high level themes (see Table 6). The most reported benefit of not developing AI was HUMAN GROWTH AND POTENTIAL (43.4%), such as less dependence on tech (26.4%), learning skills and knowledge (20.7%), and increased human interaction and dependence on one another (13.2%). The second most common benefit was ECONOMIC IMPACT AND ECONOMIC SECURITY, such as job security (17.6%) and financial benefits (6.8%). Regarding the scale of benefit (Q22, see Table 6), BENEFICIAL SIDE EFFECTS OF NOT USING AI (e.g., better health and environmental impact) had the highest impact (4.75±2.71). HUMAN GROWTH AND POTENTIAL also had a high impact (3.95±2.71), showing an tension in delaying technological progress for the sake of not LIMITING HUMAN POTENTIAL.

4.4 RQ4. Tension between Developing and Not Developing the Applications

We analyzed participants' opinions on use case development, their decision confidence, and perceived agreement from others to identify the source of tension. Most participants answered that the use case "should" (83%) rather than "should not" (17%) be developed.

Factors in Tensions over Development To examine how considering harms and benefits impacted opinions on whether a use case should be developed, we ran a linear

Code	Quote & Scale of Impact (Q13 / Q16)	% responses Q11 295×	Q14 295×
SOCIAL & PSYCHOLOGICAL EFFECT	**5.82±1.94 / 5.29±2.05**	**35.3**	**26.4**
Manipulate people	*"People would lose control potentially over important data, ideas..."* (P113, Q11)	12.9	1.0
Misinformation	*"It could give false information and confuse people as to where they don't know which source of information to trust"* (P126, Q11)	12.5	10.8
Mental harm	*"Loss of confidence and motivation: Repeated misunderstandings and failed interactions could lead to frustration and a reluctance to engage in...learning."* (P280, Q14)	12.2	9.5
Social isolation	*"It would harm...relationships...maybe [leading] to ostracism or loss of trust."* (P200, Q14)	2.4	4.4
ECONOMIC IMPACT	**5.39±1.70 / 4.63±1.75**	**32.5**	**33.6**
Financial disturbance	*"People would lose jobs and incomes"* (P93, Q11)	16.3	20.7
Economic disturbance	*"Shortage in suppliers or a raise in costs"* (P126, Q11)	12.9	9.8
Waste resources or time	*"Potentially leading to misguided decisions [and] wasted resources..."* (P250, Q14)	1.0	9.5
SAFETY & SECURITY RISK	**6.32±1.57 / 6.10±1.60**	**21.7**	**7.8**
Data security & privacy risk	*"It could be...compromising users' private data"* (P235, Q14)	10.5	3.1
Extinction	*"human race eliminated by machines"* (P220, Q11)	5.8	2.0
Aid criminal	*"It will lead to theft"* (P275, Q14)	4.1	2.7
PHYSICAL EFFECT	**6.70±1.40 / 6.28±1.59**	**16.9**	**29.8**
Physical harm	*"It could lead to serious injury or death."* (P215, Q14)	12.9	23.7
Negative health & well-being	*"People would become more unhealthy"* (P175, Q11)	3.4	8.5
QUALITY & RELIABILITY ISSUES OF AI	**5.16±1.76 / 5.43±2.12**	**15.9**	**24.1**
Incorrect AI output	*"Providing incorrect or incomplete medical diagnostics"* (P221, Q14)	9.8	13.6
Distrust AI	*"People would lose trust in technology"* (P72, Q11)	5.1	5.8
IMPEDING HUMAN DEVELOPMENT & LEARNING	**4.41±1.80 / 4.42±2.10**	**12.9**	**12.5**
Overreliance	*"[People] will not learn to do anything on their own"* (P274, Q14)	9.5	6.8
Impede learning	*"Diminished capacity for original ideas, maybe even critical thinking"* (P221, Q11)	3.7	3.7
Hinder career	*"The reputation of the developers would be ruined"* (P125, Q14)	1.0	4.1
REDUCING QUALITY & RELIABILITY OF SOCIETY	**5.70±2.01 / 5.48±2.09**	**12.2**	**9.2**
Distrust institution	*"The negative impact would be...decreased trust in the medical professionals..."* (P62, Q14)	5.8	5.4
Legal issues	*"Increased lawsuits."* (P95, Q14)	3.1	4.1
GENERAL HARM	**5.95±2.25 / 5.97±1.94**	**6.8**	**11.5**
General harm	*"More people would be hurt"* (P105, Q11)	4.7	10.2
Range	*"Could cause anything from minor issues to loss of life"* (P271, Q14)	2.0	1.7
REDUCING PROGRESS	**7.50±0.84 / 6.57±1.45**	**2.0**	**4.7**
Environmental harm	*"It could negatively impact fighting climate change"* (P23, Q14)	1.4	1.7
Hinder science	*"Delay scientific advancement and progress"* (P276, Q14)	0.7	3.1
OTHER	**3.30±3.22 / 4.28±2.83**	**4.4**	**7.1**
N/A	*"Many things"* (P216, Q11)	2.4	2.7
No harm	*"I can't think of any [harms]"* (P242, Q11)	1.7	3.7

Table 3: Harms of developing (Q11, Q14): percentage of occurrence for THEME with scale of impact (Q13, Q16) and corresponding top few most frequent codes with representative quotes.

mixed effects model with the opinion as dependent variable and the scale of benefits and harms as the independent variables (see Table 7). Surprisingly, *Harms of not developing* consistently showed the most significant effect on opinions ($\beta = 0.27$, $p < 0.001$), confidence ($\beta = 0.41$, $p < 0.001$), and agreement ($\beta = 0.29$, $p < 0.001$) that the application *should* be developed. Similarly, *benefits of not developing* showed the most significant effect on the opinion that the application *should not* be developed (opinions; $\beta = -0.20$, $p < 0.001$, confidence; $\beta = -0.23$, $p < 0.001$, agreement; $\beta = -0.22$, $p < 0.001$). These results highlight that considering *not developing* scenarios provides deeper insights into people's opinions about AI development than *developing* scenarios alone. Interestingly, despite the harms and benefits of not developing are use case specific, these answers generally reflected participants' general attitudes toward AI. Specifically, among people who believe their selected use cases should not be developed, 54.9% and 52.9% of them also think Tech-X (Q1) and Tech-X 10 (Q4) "Should not exist," respectively, higher than proportions from all the participants (13.9%; Q1, 14.6%; Q4).

Case Analysis Ordering domains and themes by the number of use cases participants said should not be developed, applications that impact WORK (28.1%) ranked first, followed by SOCIETY (21.4%) and GOAL (18.2%); see Appendix E.2 for further analysis. Examples of similar tasks that participants said should and should not be developed are shown in Table 6 (see Appendix C). P189 ("Should not be developed") and P32 ("Should be developed") both discuss reduced accessibility in the harms of not developing an AI assisted transportation system. However, they focus on different benefits of not developing where P32 states that those

Code	Quote & Scale of Impact (Q10)	% responses Q8 295×
REINVEST HUMAN CAPITAL	**5.13±1.86**	**52.5**
Personal life efficiency	*"Save time, effort, and energy...[and] allow a layperson to accomplish this task."* (P19)	35.6
Personal growth	*"Since it's data driven, individual performances will be vigorously assessed and suggest ways by which an individual can improve."* (P47)	16.9
Reduce mundane work	*"I would be able to focus on relationships and team building versus menial manager tasks that AI could complete for me."* (P157)	13.2
ECONOMIC GAIN	**5.38±1.75**	**43.7**
General efficiency	*"Companies will not need to have as many employees...because they'll be able to automate much of the workload...which will increase company profits."* (P209)	31.5
Financial gain	*"It will save cost of different diagnostic tests."* (P211)	17.6
RESOURCE ACCESSIBILITY	**5.58±1.62**	**35.9**
Information accessibility	*"People need quick and reliable answers because not a lot of people have time for themselves...[and] can't deeply engage in topics they encounter in daily life."* (P53)	18.0
Resource accessibility	*"It would give people more equity and assistance."* (P99)	15.6
IMPROVE SOCIETAL ISSUES	**6.96±1.25**	**31.9**
Improve medical care	*"Health care would be cheaper (hopefully) and more accessible to everyone"* (P109)	13.2
Scientific research innovation	*"Research would be able to be done at a faster pace."* (P159)	11.2
REDUCE ERROR	**6.11±1.57**	**16.9**
Less human error	*"It could remove certain human biases"* (P171)	10.8
Information quality	*"It could quickly detect lies said by politicians."* (P229)	9.8
IMPROVE QUALITY OF PERSONAL LIFE	**6.50±1.50**	**15.3**
Improve well-being & health	*"My family would have a healthier diet & they would live better lives."* (P238)	10.2
Improve mental health	*"It would reduce the cases of mental illness in lonely people."* (P96)	5.4
IMPROVE QUALITY OF SOCIAL LIFE	**6.76±1.17**	**8.1**
Better communication	*"People would be able to communicate in different languages in real-time."* (P176)	5.4
Social interaction	*"It would help me navigate through various social situations and problems, thus improving my social life."* (P200)	2.0
OTHER	**5.25±2.17**	**5.4**
New code	*"Help others with problems"* (P232)	2.7

Table 4: Benefits of developing (Q8): Percentage of occurrence for THEME with scale of impact (Q10) and corresponding top few most frequent codes with representative quotes.

negatively affected by not developing "*will miss out on being more independent*" whereas P189 reflects on "*human workers*" keeping their jobs. Meanwhile, P175 ("Should not be developed") and P10 ("Should be developed") discuss similar lack of assistance in addressing mundane task of buying groceries, but P10 noticed less reliance whereas P175 noticed not developing would increase human interaction by forcing people to "*go outside and interact in society*". Interestingly, we find that in answers to benefits of not developing, participants who thought the application should not be developed focused more on alternate solutions having additional benefits not addressable by technology whereas those who thought the application should be developed focused on the absence of possible harms from technology.

5 Discussion & Conclusion

To address the need for lay user participation in anticipating the harms and benefits of AI use cases, we introduced PARTICIP-AI, a framework to collect diverse AI use cases and examine the benefits and harms of both developing and *not* developing them. Using our framework, we collected AI use cases from nearly 300 demographically diverse participants. We now discuss the implications of our findings on future work towards democratic AI development and policy.

Benefits of AI: Augmenting Life and Social Good. The brainstorming exercises in our framework uncovered a new array of AI usage highlighting personal life applications to augment everyday life, and societal applications to enrich the lives of everyone. At a personal level, participants expressed interest in automating daily tasks, assisting personal growth including mental and physical health, and better allocation of resources (§4.2), echoing the need for AI design to allow greater stakeholder liberties (Bondi et al. 2021).

Participants showed strong interest in using AI to solve significant societal problems from advances in medicine to addressing inequality, global warming, and world hunger. These use cases, present a stark contrast to the current directions of AI development geared towards work and business productivity (Maslej et al. 2023). Thus, future studies should explore methods to satisfy these public needs like digital commons (Verdegem 2022), moving beyond profitability.

Harms Envisioned by Lay Users. Our framework also enabled participants to reason through harms and benefits of AI use cases, where a unique set of harms emerged. Our results show that lay users can anticipate the impacts of AI in their daily lives, complementary to technical experts' assessment (Solaiman et al. 2023; Weidinger et al. 2023). While the themes had some overlap with Solaiman et al. (2023), ad-

Code	Quote & Scale of Impact (Q19)	% responses Q17 295×
LIMITING HUMAN POTENTIAL	**3.56±2.12**	**32.9**
Waste resources or time	*"I waste so much time on these types of activities. Time that could be spent on productive things..."* (P39)	13.2
Inefficiency	*"Government and public agencies will continue to operate in a wasteful and ineffective manner."* (P76)	12.2
Impede personal growth	*"People would not be able to reach their potentials."* (P106)	11.9
LOSE INFORMATION & ACCESSIBILITY TO RESOURCES	**4.14±2.04**	**26.1**
Lose assistance	*"Immigrants would not recieve [sic] translation support easily."* (P122)	11.5
Lose solution or service	*"Homeless need easier more accessible help..."* (P206)	9.8
OTHER	**2.73±2.59**	**23.4**
No harm	*"It wouldnt [sic] necessarily be harmful"* (P216)	16.6
New code	*"It may be an emergency"* (P1)	4.1
LESS INNOVATION	**4.64±2.08**	**15.9**
Delay in innovation	*"It could slow down progress against climate change"* (P23)	9.5
Less innovation	*"New technology would not be used to help man kind."* (P87)	8.5
SOCIAL & PSYCHOLOGICAL EFFECT	**4.07±2.40**	**15.6.5**
Stress & overworked	*"It would increase the workload and time spent on tedious tasks."* (P181)	10.8
Mental harm	*"It would deprive people of an opportunity to address their loneliness"* (P96)	4.7
LESS PROGRESS IN SOLVING SOCIETAL ISSUES	**4.78±2.19**	**13.9.5**
Hinder medical care	*"Many individuals will continue suffering from ailments that...worsen in time."* (P117)	8.5
Misinformation	*"Some people...find bad answers on the internet that make things worse"* (P246)	3.7
ECONOMIC & BUSINESS IMPACT	**3.70±2.27**	**11.5.5**
Financial disturbance	*"Individuals might lack access to highly personalized and good retirement strategies"* (P292)	4.7
Economic disturbance	*"Increases cost and reduce employment"* (P16)	4.1
LIMITED TO HUMAN CAPABILITIES	**4.39±1.64**	**9.5**
Human error	*"Humans are biased...and often unable to combine various fields of thought."* (P88)	6.8
Hinder creative work	*"It could hinder some people's ability to create."* (P120)	2.7
PHYSICAL EFFECT	**5.14±2.62**	**8.5**
Physical harm	*"it could've saved a lot of lives"* (P152)	5.8
Health issues	*"My health will suffer."* (P30)	3.7

Table 5: Harms of *not* developing (Q17): percentage of occurrence for THEME with scale of impact (Q19) and corresponding top few most frequent codes with representative quotes.

ditional or more detailed harms were uncovered such as distrust in AI and technology, distrust in institutions, and stalled progress (§4.2). Experts also discussed harms that were not as common in participant responses such as environmental costs, and data labor, highlighting the complementary value of the approaches. We project that AI literacy could further empower non-expert public to surface and discuss more diverse and relevant harms (Long and Magerko 2020).

Psychological harms such as manipulation, misinformation, and mental harm were among the most common concerns (§4.2), however, have been largely overlooked in current regulatory and academic discussions on AI. Some emerging works have examined the psychological impact of AI and automation (e.g., depression (Vidal et al. 2020; McClure 2022), influence on autonomy (Hackenburg and Margetts 2023; Jakesch et al. 2023), over-reliance (Ma, Mei, and Su 2023)); but the negative impacts of AI remains largely under-explored (Farahany 2023). Concerningly, these intangible yet impactful harms would not be effectively remedied through law and policy like EU AI Act (Pałka 2023) or US liability case law (Cheong, Caliskan, and Kohno 2023). Therefore, further studies to understand how AI affects mental health are paramount to establishing frameworks that can reveal harms to hold different actors accountable.

Techno-solutionism and Tensions of (Not) Developing. As seen with many examples (Haven and Boyd 2020), AI applied without careful consideration can exacerbate the existing inequalities by creating a hierarchy of the technology owner and the recipient, especially through its opaqueness (Madianou 2021). In addition to the harms of AI such as disparate performance on majority vs. minority groups (Buolamwini and Gebru 2018; Sap et al. 2019; Scheuerman, Paul, and Brubaker 2019), risk of dual-use (Kaffee et al. 2023), and imposition of norms (Santy et al. 2023), the foregone benefits of non-technical solutions such as job creation, human involvement, and community building should be further studied and considered when discussing risks and harms of AI, as illustrated by our framework.

By collecting and analyzing harms and benefits under two alternate scenarios (to develop and not develop), we aimed to understand the reasoning behind users' decisions. Our qualitative analyses showed that participants often emphasized the benefits of non-technical solutions, such as increased social interaction, job security, and positive impacts toward indirect stakeholders, when they opted for not developing the use case (§4.4). This highlights the need for discussions of

Code	Quote & Scale of Impact (Q22)	% responses Q20 295×
HUMAN GROWTH & POTENTIAL	**3.95±2.08**	**43.4**
Less dependent on technology	*"People would think critically and rely on the thoughts of other human beings who have a more nuanced understanding of real life situations then AI ever could."* (P113)	26.4
Learning skills & knowledge	*"It would make it so more people would strive to learn the local language"* (P122)	20.7
Human interaction dependence	*"I might have to communicate that I need help and hopefully would bring us together."* (P249)	13.2
ECONOMIC IMPACT / SECURITY	**3.54±2.23**	**22.4**
Job security	*"It will not take over peoples' jobs."* (P251)	17.6
Financial benefit	*"The insurance companies and doctors...make more money off of multiple visits"* (P272)	6.8
OTHER	**3.05±2.79**	**18.6**
No benefit	*"I don't see any benefits"* (P272)	10.8
New code	*"The company could put in more effort"* (P148)	4.4
LESS BAD AI USAGE	**3.66±2.41**	**15.6**
Less improper or unethical use	*"It would not allow for the potential harmful uses of the ai assistance"* (P235)	9.5
More privacy	*"It would protect information of all and keep breaches at a minimum"* (P293)	4.4
INCREASE TRANSPARENCY, CONTROL, & RELIABILITY	**3.34±2.06**	**12.5**
More attentive	*"It could foster more personal involvement...in one's investment choices"* (P31)	7.1
Human control	*"It allows for more deliberate, controlled, and transparent progress...fostering public trust and the responsible development of technology."* (P204)	4.1
NO CHANGES	**3.26±2.33**	**7.8**
Maintain status quo	*"There would really be no change in society, it would remain the same"* (P216)	5.8
Other non-AI solutions	*"Advances in medicine woud still occur with use of other technologies and methods"* (P45)	2.0
BENEFICIAL SIDE EFFECTS OF NOT USING AI	**4.75±2.71**	**2.7**
Better health	*"People...[would]...seek qualified medical assistance, which could save their life."* (P43)	2.0
Environmental	*"AI requires a lot of energy so not developing it will be good for the environment"* (P209)	0.7

Table 6: Benefits of *not* developing (Q20): percentage of occurrence for THEME with scale of impact (Q22) and corresponding top few most frequent codes with representative quotes.

	Development Opinion (Q23)		Confidence (Q24)		Agreement (Q25)	
	Coefficient (SE)	p-value	Coefficient (SE)	p-value	Coefficient (SE)	p-value
Benefits of Developing (Q10)	0.16 (0.07)	$< .05$	0.17 (0.06)	$< .01$	0.20 (0.06)	$< .01$
Harms of Developing (Q13)	−0.08 (0.07)	0.30	−0.12 (0.06)	$< .05$	−0.10 (0.07)	0.19
Harms of Developing (Q15)	−0.12 (0.07)	0.08	−0.18 (0.07)	$< .01$	−0.08 (0.07)	0.22
Harms of Not Developing (Q19)	**0.27** (0.07)	$< .001$	**0.41** (0.06)	$< .001$	**0.29** (0.06)	$< .001$
Benefits of Not Developing (Q22)	**−0.20** (0.06)	$< .001$	**−0.23** (0.05)	$< .001$	**−0.22** (0.06)	$< .001$

Table 7: Effects of each benefit and harms scale to the development opinion {-1, 1}, confidence {-4, 4}, and agreement {-8, 8}. All scales were normalized, and negative values denote opinions "Should not be developed". Standard error is in parenthesis.

often overlooked non-technical solutions and their benefits to various stakeholders, particularly to those vulnerable and marginalized, beyond the default persona of technology (i.e., a culturally prototypical user, often straight white tech-savvy men) (Sim et al. 2023). Thus, anticipation of consequences (Do et al. 2023) in making development decisions could be a promising direction towards an inclusive progress.

Participants' responses on harms and benefits of not developing the AI system also highlighted tensions around human growth and potential. Not developing a use case could reduce the efficiency of allocating human resources, but the absence of AI applications could fortify human worth and independence, spurring investment in human knowledge and skills. This dilemma underscores the tension between human's value in creation activities and its perceived competition with that of the machines. This resonates with creator groups' call for protective regulations for their work (aut 2023; The Authors Guild 2024) and researchers' warning against greater inequality from AI-induced productivity (Brynjolfsson 2023; Littman et al. 2022; Moradi and Levy 2020; Cheong, Caliskan, and Kohno 2023). Given these concerns, researchers, developers, and companies should consider immediate and long term impacts of AI in labor to maintain the value of human work. In developing AI, a focus on implementing participatory approaches to ensure positive and mitigate negative impacts on affected communities (Bondi et al. 2021; Delgado et al. 2023b; Floridi et al. 2021). Additionally, regulatory measures and economic policies must aim to ensure human value and equality in the distribution of AI-generated benefits (Littman et al. 2022).

Ethical Considerations

Our research endeavor, while aimed at inclusivity, predominantly involved participants from the United States who are English speakers, a demographic feature that carries significant ethical implications regarding the generalizability and inclusivity of our findings (Lee and Rich 2021). We recognize the potential marginalization of non-English speakers and individuals outside of the U.S., and the ethical responsibility to ensure that the frameworks we investigate are adaptable across diverse cultural and linguistic contexts. This calls for future research to extend our framework to a broader array of participants, including those from varied cultural backgrounds and those less familiar with technology. Doing so would not only enhance the robustness of our findings but also uphold the ethical principle of inclusivity.

Moreover, the use of an online survey platform, such as Prolific, inherently skews our sample towards individuals who are more comfortable with and have access to technology. This presents an ethical consideration regarding the digital divide and the potential exclusion of technologically underserved populations. Ethical research practice necessitates actively seeking ways to include these populations to avoid reinforcing existing disparities.

Additionally, while we manually verified the quality of our study's data, crowdsourcing-based studies are decentralized and difficult to guarantee the reliability of the data. Moreover, in our analysis, we used GPT-4 to code our data; therefore, with our released data, future work could explore other approaches such as clustering, fully manual coding, coding based on predefined taxonomy, or improving GPT-4's coding abilities. The survey wording, formatting and ordering could have affected the participant answers (McFARLAND 1981), and future works should further explore effects of ordering and phrasing of the questions and descriptions of the future technologies.

Acknowledgements

We thank the Prolific workers for their thoughtful answers. This work was supported, in whole or in part, by the Block Center for Technology and Society at Carnegie Mellon University, in part by DARPA under the ITM program (FA8650-23-C-7316), and by the University of Washington Tech Policy Lab, which receives support from the William and Flora Hewlett Foundation, the John D. and Catherine T. MacArthur Foundation, Microsoft, and the Pierre and Pamela Omidyar Fund at the Silicon Valley Community Foundation. Additionally, Jenny T. Liang was supported by the National Science Foundation under grants DGE1745016 and DGE2140739.

References

2023. Authors Guild v. OpenAI Inc., 23-cv-8292 (SHS) (S.D.N.Y. Nov. 29).

Anthropic, C. I. P. 2023. Collective Constitutional AI: Aligning a Language Model with Public Input. *anthropic.com*.

Bai, Y.; Jones, A.; Ndousse, K.; Askell, A.; Chen, A.; DasSarma, N.; Drain, D.; Fort, S.; Ganguli, D.; Henighan, T.; et al. 2022a. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.

Bai, Y.; Kadavath, S.; Kundu, S.; Askell, A.; Kernion, J.; Jones, A.; Chen, A.; Goldie, A.; Mirhoseini, A.; McKinnon, C.; Chen, C.; Olsson, C.; Olah, C.; Hernandez, D.; Drain, D.; Ganguli, D.; Li, D.; Tran-Johnson, E.; Perez, E.; Kerr, J.; Mueller, J.; Ladish, J.; Landau, J.; Ndousse, K.; Lukosuite, K.; Lovitt, L.; Sellitto, M.; Elhage, N.; Schiefer, N.; Mercado, N.; DasSarma, N.; Lasenby, R.; Larson, R.; Ringer, S.; Johnston, S.; Kravec, S.; Showk, S. E.; Fort, S.; Lanham, T.; Telleen-Lawton, T.; Conerly, T.; Henighan, T.; Hume, T.; Bowman, S. R.; Hatfield-Dodds, Z.; Mann, B.; Amodei, D.; Joseph, N.; McCandlish, S.; Brown, T.; and Kaplan, J. 2022b. Constitutional AI: Harmlessness from AI Feedback. arXiv:2212.08073.

Balaram, B.; Greenham, T.; and Leonard, J. 2018. Artificial Intelligence: real public engagement. *RSA, London. Retrieved November*, 5: 2018.

Barrett, A. M.; Newman, J.; Nonnecke, B.; Hendrycks, D.; Murphy, E. R.; and Jackson, K. 2023. AI risk-management standards profile for general-purpose AI systems (GPAIS) and foundation models. *Center for Long-Term Cybersecurity, UC Berkeley. https://perma. cc/8W6P-2UUK*.

Baumann, K.; Caldwell, B.; Bar, F.; and Stokes, B. 2018. Participatory design fiction: community storytelling for speculative urban technologies. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–1.

Baumann, K.; Stokes, B.; Bar, F.; and Caldwell, B. 2016. Designing in" constellations" sustaining participatory design for neighborhoods. In *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops-Volume 2*, 5–8.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.

Bengio, Y.; Hinton, G.; Yao, A.; Song, D.; Abbeel, P.; Harari, Y. N.; Zhang, Y.-Q.; Xue, L.; Shalev-Shwartz, S.; Hadfield, G.; Clune, J.; Maharaj, T.; Hutter, F.; Baydin, A. G.; McIlraith, S.; Gao, Q.; Acharya, A.; Krueger, D.; Dragan, A.; Torr, P.; Russell, S.; Kahneman, D.; Brauner, J.; and Mindermann, S. 2023. Managing AI Risks in an Era of Rapid Progress. arXiv:2310.17688.

Bergman, A. S.; Hendricks, L. A.; Rauh, M.; Wu, B.; Agnew, W.; Kunesch, M.; Duan, I.; Gabriel, I.; and Isaac, W. 2023. Representation in AI Evaluations. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 519–533.

Birhane, A.; Isaac, W.; Prabhakaran, V.; Diaz, M.; Elish, M. C.; Gabriel, I.; and Mohamed, S. 2022. Power to the people? opportunities and challenges for participatory AI. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–8.

Bleecker, J. 2022. *Design Fiction*, chapter 24, 561–578. John Wiley & Sons, Ltd. ISBN 9781119815075.

Bondi, E.; Xu, L.; Acosta-Navas, D.; and Killian, J. A. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 425–436. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8473-5.

Bostyn, D. H.; Sevenhant, S.; and Roets, A. 2018. Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas. *Psychological Science*, 29(7): 1084–1093. PMID: 29741993.

Brynjolfsson, E. 2023. The turing trap: The promise & peril of human-like artificial intelligence. In *Augmented Education in the Global Age*, 103–116. Routledge.

Buçinca, Z.; Pham, C. M.; Jakesch, M.; Ribeiro, M. T.; Olteanu, A.; and Amershi, S. 2023. Aha!: Facilitating ai impact assessment by generating examples of harms. *arXiv preprint arXiv:2306.03280*.

Buolamwini, J.; and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Friedler, S. A.; and Wilson, C., eds., *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, 77–91. PMLR.

Card, D.; and Smith, N. A. 2020. On consequentialism and fairness. *Frontiers in Artificial Intelligence*, 3: 34.

Center, P. R. 2023. Public Awareness of Artificial Intelligence in Everyday Activities.

Cheong, I.; Caliskan, A.; and Kohno, T. 2023. Is the U.S. Legal System Ready for AI's Challenges to Human Values? arXiv:2308.15906.

Christiano, P.; Leike, J.; Brown, T. B.; Martic, M.; Legg, S.; and Amodei, D. 2023. Deep reinforcement learning from human preferences. arXiv:1706.03741.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1): 37–46.

Corbett, E.; Denton, E.; and Erete, S. 2023. Power and public participation in ai. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–13.

Costanza-Chock, S. 2018. Design justice: Towards an intersectional feminist framework for design theory and practice. *Proceedings of the Design Research Society*.

Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 2023a. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–23.

Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 2023b. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703812.

Dindler, C.; and Iversen, O. S. 2007. Fictional Inquiry—design collaboration in a shared narrative space. *CoDesign*, 3(4): 213–234.

Do, K.; Pang, R. Y.; Jiang, J.; and Reinecke, K. 2023. "That's important, but...": How computer science researchers anticipate unintended consequences of their research innovations. In *CHI*, volume 1. Association for Computing Machinery.

Durmus, E.; Nyugen, K.; Liao, T. I.; Schiefer, N.; Askell, A.; Bakhtin, A.; Chen, C.; Hatfield-Dodds, Z.; Hernandez, D.; Joseph, N.; et al. 2023. Towards measuring the representation of subjective global opinions in language models. *arXiv preprint arXiv:2306.16388*.

Evtimov, I.; Cui, W.; Kamar, E.; Kiciman, E.; Kohno, T.; and Li, J. 2020. Security and Machine Learning in the Real World. arxiv:2007.07205.

Farahany, N. A. 2023. *The battle for your brain: defending the right to think freely in the age of neurotechnology*. St. Martin's Press.

Farias, P. G.; Bendor, R.; and Van Eekelen, B. F. 2022. Social dreaming together: A critical exploration of participatory speculative design. In *Proceedings of the Participatory Design Conference 2022-Volume 2*, 147–154.

Floridi, L.; Cowls, J.; King, T. C.; and Taddeo, M. 2021. How to design AI for social good: seven essential factors. *Ethics, Governance, and Policies in Artificial Intelligence*, 125–151.

Forlano, L. E.; and Halpern, M. K. 2023. Speculative histories, just futures: From counterfactual artifacts to counterfactual actions. *ACM Transactions on Computer-Human Interaction*, 30(2): 1–37.

Friedman, B.; and Hendry, D. 2012. The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1145–1148.

Friedman, B.; Kahn Jr., P. H.; and Borning, A. 2008. *Value Sensitive Design and Information Systems*, chapter 4, 69–101. John Wiley & Sons, Ltd. ISBN 9780470281819.

Fry, T. 2009. Design futuring. *University of New South Wales Press, Sydney*, 71–77.

Hackenburg, K.; and Margetts, H. 2023. Evaluating the persuasive influence of political microtargeting with large language models.

Haven, J.; and Boyd, D. 2020. Philanthropy's technosolutionism problem. *Data & Society Research Institute*.

Hohendanner, M.; Ullstein, C.; Buchmeier, Y.; and Grossklags, J. 2023. Exploring the Reflective Space of AI Narratives Through Speculative Design in Japan and Germany. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 351–362.

Jakesch, M.; Bhat, A.; Buschek, D.; Zalmanson, L.; and Naaman, M. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15.

Kaffee, L.-A.; Arora, A.; Talat, Z.; and Augenstein, I. 2023. Thorny Roses: Investigating the Dual Use Dilemma in Natural Language Processing. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 13977–13998. Singapore: Association for Computational Linguistics.

Kohno, T.; Acar, Y.; and Loh, W. 2023. Ethical Frameworks and Computer Security Trolley Problems: Foundations for Conversations. In *32nd USENIX Security Symposium (USENIX Security 23)*, 5145–5162. ISBN 978-1-939133-37-3.

Kozubaev, S.; Elsden, C.; Howell, N.; Søndergaard, M. L. J.; Merrill, N.; Schulte, B.; and Wong, R. Y. 2020. Expanding modes of reflection in design futuring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Lee, M. K.; and Rich, K. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Littman, M. L.; Ajunwa, I.; Berger, G.; Boutilier, C.; Currie, M.; Doshi-Velez, F.; Hadfield, G.; Horowitz, M. C.; Isbell, C.; Kitano, H.; Levy, K.; Lyons, T.; Mitchell, M.; Shah, J.; Sloman, S.; Vallor, S.; and Walsh, T. 2022. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. arXiv:2210.15767.

Long, D.; and Magerko, B. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Lubars, B.; and Tan, C. 2019. *Ask Not What AI Can Do, but What AI Should Do: Towards a Framework of Task Delegability*. Red Hook, NY, USA: Curran Associates Inc.

Ma, Z.; Mei, Y.; and Su, Z. 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings*, volume 2023, 1105. American Medical Informatics Association.

Madianou, M. 2021. Nonhuman humanitarianism: when 'AI for good' can be harmful. *Information, Communication & Society*, 24(6): 850–868.

Maslej, N.; Fattorini, L.; Brynjolfsson, E.; Etchemendy, J.; Ligett, K.; Lyons, T.; Manyika, J.; Ngo, H.; Niebles, J. C.; Parli, V.; Shoham, Y.; Wald, R.; Clark, J.; and Perrault, R. 2023. The AI Index 2023 Annual Report. Technical report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.

Matter, D.; Schirmer, M.; Grinberg, N.; and Pfeffer, J. 2024. Close to Human-Level Agreement: Tracing Journeys of Violent Speech in Incel Posts with GPT-4-Enhanced Annotations. *arXiv preprint arXiv:2401.02001*.

McClure, K. 2022. Case for Protecting Youth from the Harmful Mental Effects of Social Media. *Chapman Law Review*, 26(1): 325–360.

McFARLAND, S. G. 1981. Effects of Question Order on Survey Responses. *Public Opinion Quarterly*, 45(2): 208–215.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

Miceli, M.; and Posada, J. 2022. The Data-Production Dispositif. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–37.

Moradi, P.; and Levy, K. 2020. The Future of Work in the Age of AI. In *The Oxford Handbook of Ethics of AI*, 269–288. Oxford University Press Oxford.

Nekoto, W.; Marivate, V.; Matsila, T.; Fasubaa, T.; Kolawole, T.; Fagbohungbe, T.; Akinola, S. O.; Muhammad, S. H.; Kabongo, S.; Osei, S.; et al. 2020. Participatory research for low-resourced machine translation: A case study in african languages. *arXiv preprint arXiv:2010.02353*.

NIST. 2023. Biden-Harris Administration Announces New NIST Public Working Group on AI.

OpenAI. 2023. Our approach to AI safety. Available at https://openai.com/blog/our-approach-to-ai-safety (2024/01/10).

Parliament, E. 2023. Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI.

Pasquale, F.; Malgieri, G.; and Pasquale, M. 2021. If you don't trust AI yet, you're not wrong. *The New York Times*.

Pałka, P. 2023. *AI, Consumers & Psychological Harm*. Cambridge University Press. Forthcoming.

Pöhler, L.; Schrader, V.; Ladwein, A.; and von Keller, F. 2024. A Technological Perspective on Misuse of Available AI. *arXiv preprint arXiv:2403.15325*.

Prabhakaran, V.; Mitchell, M.; Gebru, T.; and Gabriel, I. 2022. A human rights-based approach to responsible AI. *arXiv preprint arXiv:2210.02667*.

Queerinai, O. O.; Ovalle, A.; Subramonian, A.; Singh, A.; Voelcker, C.; Sutherland, D. J.; Locatelli, D.; Breznik, E.; Klubicka, F.; Yuan, H.; et al. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1882–1895.

Raji, I. D.; Kumar, I. E.; Horowitz, A.; and Selbst, A. 2022. The fallacy of AI functionality. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 959–972.

Rao, A.; and Kreps, S. 2023. How to systemically think about AI regulation. https://www.brookings.edu/articles/how-to-systemically-think-about-ai-regulation/. Accessed: 2024-1-22.

Reeve, O.; Colom, A.; and Modhvadia, R. 2023. What Do the Public Think about AI? https://www.adalovelaceinstitute.org/evidence-review/what-do-the-public-think-about-ai/.

Saldaña, J. 2021. The coding manual for qualitative researchers. *The coding manual for qualitative researchers*, 1–440.

Santy, S.; Liang, J. T.; Le Bras, R.; Reinecke, K.; and Sap, M. 2023. NLPositionality: Characterizing Design Biases of Datasets and Models. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Sap, M.; Card, D.; Gabriel, S.; Choi, Y.; and Smith, N. A. 2019. The Risk of Racial Bias in Hate Speech Detection. In Korhonen, A.; Traum, D.; and Màrquez, L., eds., *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678. Florence, Italy: Association for Computational Linguistics.

Scheuerman, M. K.; Paul, J. M.; and Brubaker, J. R. 2019. How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).

Sim, M.; Hugenberg, K.; Kohno, T.; and Roesner, F. 2023. A Scalable Inclusive Security Intervention to Center Marginalized & Vulnerable Populations in Security & Privacy Design. In *Proceedings of the 2023 New Security Paradigms Workshop*, NSPW '23, 102–115. New York, NY, USA: Association for Computing Machinery. ISBN 9798400716201.

Solaiman, I.; Talat, Z.; Agnew, W.; Ahmad, L.; Baker, D.; Blodgett, S. L.; au2, H. D. I.; Dodge, J.; Evans, E.; Hooker, S.; Jernite, Y.; Luccioni, A. S.; Lusoli, A.; Mitchell, M.; Newman, J.; Png, M.-T.; Strait, A.; and Vassilev, A. 2023. Evaluating the Social Impact of Generative AI Systems in Systems and Society. arXiv:2306.05949.

Suresh, H.; Movva, R.; Dogan, A. L.; Bhargava, R.; Cruxên, I.; Cuba, Á. M.; Taurino, G.; So, W.; and D'Ignazio, C. 2022. Towards intersectional feminist and participatory ML: A case study in supporting Feminicide Counterdata Collection. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 667–678.

The Authors Guild. 2024. Artificial Intelligence. https://authorsguild.org/advocacy/artificial-intelligence/. Accessed 09-01-2024.

Tracy, R. 2023. ChatGPT's Sam Altman Warns Congress That AI 'Can Go Quite Wrong'. *Wall Street Journal*. Accessed 09-01-2024.

Trustible. 2023. A Machine Learning Engineer's Guide To The AI Act. *Forbes Magazine*.

van der Veer, S. N.; Riste, L.; Cheraghi-Sohi, S.; Phipps, D. L.; Tully, M. P.; Bozentko, K.; Atwood, S.; Hubbard, A.; Wiper, C.; Oswald, M.; et al. 2021. Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries. *Journal of the American Medical Informatics Association*, 28(10): 2128–2138.

Verdegem, P. 2022. Dismantling AI capitalism: the commons as an alternative to the power concentration of Big Tech. *AI & SOCIETY*.

Vidal, C.; Lhaksampa, T.; Miller, L.; and Platt, R. 2020. Social Media Use and Depression in Adolescents: A Scoping Review. *International Review of Psychiatry*, 32(3): 235–253.

Viswanathan, V.; Zhao, C.; Bertsch, A.; Wu, T.; and Neubig, G. 2023. Prompt2Model: Generating Deployable Models from Natural Language Instructions. In Feng, Y.; and Lefever, E., eds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 413–421. Singapore: Association for Computational Linguistics.

Wang, Z. J.; Kulkarni, C.; Wilcox, L.; Terry, M.; and Madaio, M. 2024. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. *arXiv preprint arXiv:2402.15350*.

Weidinger, L.; Rauh, M.; Marchal, N.; Manzini, A.; Hendricks, L. A.; Mateos-Garcia, J.; Bergman, S.; Kay, J.; Griffin, C.; Bariach, B.; Gabriel, I.; Rieser, V.; and Isaac, W. 2023. Sociotechnical Safety Evaluation of Generative AI Systems. arXiv:2310.11986.

Widder, D. G.; West, S.; and Whittaker, M. 2023. Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. Available at SSRN: https://ssrn.com/abstract=4543807 or http://dx.doi.org/10.2139/ssrn.4543807.

Xiao, Z.; Yuan, X.; Liao, Q. V.; Abdelghani, R.; and Oudeyer, P.-Y. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, 75–78.

Zoph, B.; Raffel, C.; Schuurmans, D.; Yogatama, D.; Zhou, D.; Metzler, D.; Chi, E. H.; Wei, J.; Dean, J.; Fedus, L. B.; Bosma, M. P.; Vinyals, O.; Liang, P.; Borgeaud, S.; Hashimoto, T. B.; and Tay, Y. 2022. Emergent abilities of large language models. *TMLR*.